A decision tree method for explaining household gas consumption: The role of building characteristics, socio-demographic variables, psychological factors and household behaviour

Maliheh Namazkhan [a,*], Casper Albers [a], Linda Steg [b]

[a] *Department of Psychometrics and Statistics, University of Groningen, Groningen, Netherlands*
[b] *Department of Psychology, University of Groningen, Groningen, Netherlands*

A B S T R A C T

This research aims to develop a decision tree model for understanding actual gas consumption in residential buildings. Extending previous studies, this study examined to what extent four different type of factors, building characteristics, socio-demographics, psychological factors and household behaviour can explain actual gas consumption of Dutch households in 2017 and 2018. Data were collected from 601 households. A novel approach, a decision tree method, revealed that household gas consumption was related to building characteristics, socio-demographics, and psychological factors, while energy-related behaviour in households was not uniquely related to gas consumption. Specifically, house size, building age and residence type (building characteristics), household income and employment status (socio-demographics), and most notably egoistic values, hedonic values, environmental self-identity, perceived corporate environmental responsibility of the energy provider, and social norm (psychological factors) predicted total actual household gas consumption. These results illustrate that the novel integrated framework introduced in the paper yields a better understanding of actual household gas consumption. The results have important practical implications and suggest that it would be important to target these three type of factors in policy aimed to reduce household gas consumption.

1. Introduction

Energy consumption represents a major contributor to climate change and global environmental problems. Natural gas usage causes environmental problems including global warming caused by the emission of greenhouse gases. In the Netherlands, natural gas is one of the primary energy sources, making up to 41% of the total energy consumption [1]. Gas consumption is responsible for 75% of the total energy consumption in Dutch household. Gas is mostly used for heating dwellings, which made up 80% of the total household gas consumption [2]. Household gas consumption reduced by 28% between 1990 and 2018 in the Netherlands, as houses are better insulated [3]. This is a major part of household energy consumption. Besides, it is important to better understand which factors affect household gas consumption, as this reveals which strategies could be effective in reducing gas consumption and related CO_2 emissions that cause global warming and

climate change [4]; interventions will be more effective when they target key antecedents of gas use.

The aim of this paper is to study which factors are related to household gas consumption. In doing so, a wide range of factors is considered that have typically been studied in isolation. Specifically, extending previous research, this paper aims to study the relative importance of four types of factors that may affect actual household gas consumption: building characteristics, socio-demographic variables, gas use behaviour and psychological factors. These factors are explained in more detail below.

Regarding the first type of factor, building characteristics, studies have shown that residence type affects household gas consumption. For example, Dutch studies found that those living in detached house and semi-detached house used more natural gas than those living in other type of dwellings [5,6]. A study in Ireland showed that households living in detached houses use more gas than those in semi-detached houses,

probably because detached houses have more external walls and more heat loss [7]. Other studies also showed that semi-detached houses have higher gas consumption levels [8].

Furthermore, building age affects gas consumption. Some studies found that residential gas consumption was higher in older buildings, probably because of lower energy efficiency level of older buildings [9–11]. Similarly, gas consumption is higher is houses built before 1981 compared to houses built between 1981 and 2000, which could be related to newer houses having higher energy efficiency standards [7]. Another study reported that houses built between 1965 and 1974 have a higher gas consumption than older houses, which could be caused by the poor building envelope's thermal resistance of houses built in that period [6].

Regarding the impact of dwelling size on household gas consumption, studies found that a larger floor area in dwellings is associated with higher energy consumption for heating [10,12]. However, another study showed that although a larger floor area is related to higher levels of actual gas consumption in Dutch dwellings, the gas consumption per square meter is lower in larger houses [8]. Moreover, there is a positive relationship between the number of rooms heated and gas used for heating [10].

The second relevant type of factors that can influence household gas consumption are socio-demographic variables. In general, larger households use more gas [8,13,14]. In Dutch dwellings, family households use most gas while single households have the lowest gas consumption, and household without children use less gas than those with children [15].

One study showed that older people are likely to use more gas than younger respondents, probably because older people have higher comfort standards and spend more time at home [7,8]. Likewise, a study in the Netherlands, Norway, UK and Sweden reported that elderly people spend more time at home, resulting in a higher gas consumption [16].

Income is another relevant predictor of household gas consumption. Some studies reported a strong positive correlation between income and household gas consumption, presumably because higher income households live in larger dwellings [11,13,14]. In contrast, one study reported that household with lower incomes consume more gas than those with higher incomes, probably because higher income household are less often present in their home [15].

Another relevant socio-demographic variable is education level of household members [17]. Most studies found that there is a positive relationship between household education level and gas consumption [7,18,19]. This may be because households with a higher education level have a higher income and are likely to live in larger dwellings, which is associated with higher gas use.

Studies showed that household with at least one member employed used less gas than those in which no one is employed, probably because the house is less occupied when someone works [15]. Yet, not surprisingly, households who work from home are more likely to use more gas [12]. Moreover, self-employed households are likely to consume more gas than employed households, probably because self-employed households tend to earn more money than employees, and thus have more resource to afford to use more gas [7].

Third, there are many different behaviours that affect gas consumption. Particularly, as space heating accounts for a substantial proportion of overall household gas use, indoor temperature settings is a relevant behaviour that might affect household gas consumption. Indeed, studies have found that a higher average indoor temperature setting resulted in a higher energy use for space-heating [8,10].

Little is known about the relationship between psychological factors, the fourth type of factor we consider, and household gas use. It is likely that household gas consumption is influenced by various psychological factors, and particularly motivational factors. One relevant type of motivational factor are values that reflect general desirable and *trans-situational* goals, varying in importance that serve as guiding principle in individual's life [20]. Four type of values are particularly related to

environmental behaviour, including energy use [21–23]: egoistic values (i.e. focusing on increasing one's personal resources), altruistic values (i.e. reflecting concern for other human beings), hedonic values (i.e. focusing on pleasure and comfort), and biospheric values (i.e. focusing on valuing the environment). These four types of values appear to affect a range of environmental behaviours, and are therefore also likely to affect household gas consumption. Generally, people with stronger altruistic and particularly biospheric values are more likely to engage in pro-environmental behaviour. Therefore, it is likely that strong biospheric and altruistic values are related to lower levels of gas use (cf [22,24]). Strong hedonic and egoistic values are often negatively related to pro-environmental behaviours, possibly because such behaviours can be associated with less comfort and more costs [25,26]. Therefore, it is likely that strong hedonic and egoistic values are associated with higher levels of gas use.

Another relevant type of psychological factor is environmental self-identity that reflects the extent to which someone perceives himself or herself as the type of person who acts pro-environmentally [27–29]. Specifically, people with a stronger environmental self-identity are more likely to see themselves as the type of person who engages in pro-environmental actions and consequently will be more likely to act pro-environmental as people are motivated to act in line with how they see themselves. Research found that environmental self-identity is indeed related to a lower energy use [27,30–32]. Therefore, it is expected that a stronger environmental self-identity will be associated with lower levels of gas use.

Another relevant motivational factor that can affect gas use is personal norm that reflects the extent to which one feels a sense of personal obligation to engage in a certain behaviour [33]. Studies have shown that stronger personal norm to act pro-environmentally encourages pro-environmental actions [34,35]. Therefore, it is expected that a stronger personal norm to save energy is related to lower levels of gas use.

Social norm is another relevant type of motivational factor that can affect energy use [36,37]. Social norm includes injunctive norms, reflecting perception of what most people approve or disapprove, and descriptive norms, reflecting perceptions of what most other people do [38]. People are motivated to act in line with injunctive norms to get social approval or to prevent social sanctions, while they are motivated to act in line with descriptive norms as they think that what most people do is probably the most sensible thing to do [38]. It is expected that people who think that others try to reduce their gas use or expect them to reduce their gas use will use less gas.

Another relevant type of motivation factor is perceived corporate environmental responsibility of one's energy provider. Corporate environmental responsibility implies that an organisation (in this case the energy provider) has the goal to enhance its environmental performance and reduce its environmental impact [39]. When people think that their energy provider is committed to reduce its environmental impact and implemented procedures to achieve this goal, they may be more likely to think it is important to protect the environment, and they may be more motivated to engage in pro-environmental actions at home themselves. Therefore, it is expected that the more people think their energy provider endorses corporate environmental responsibility, the less gas they will use.

The aim of the research presented in this paper is to examine to what extent the factors described above are related to actual household gas use in the Netherlands. This paper extends previous research in three ways. First, as yet, most studies have focused on understanding electricity consumption. Far less studies aimed to understand household gas consumption, and in particular actual gas consumption, while this is a major energy source, particularly in the Netherlands. Second, this paper aims to study how a wide range of factors that are likely to affect household gas use, including building characteristics, socio-demographics, psychological variables, and household gas use behaviour. Notably, until now, studies typically focus on limited set of

variables, so little is known about the relative importance of these four types of factors in explaining actual household gas consumption. Specially, the effect of psychological variables on gas consumption has been understudied, while these may be relevant, as explained above. To address this gap in the literature, this study aims to test an integrated framework to enhance the understanding of factors influencing household gas consumption. Third, a novel methodological approach is used to examine to what extent these factors are related to household gas consumption: a decision tree method. A decision tree method is as a data-driven approach that classifies different groups of households on the basis of the characteristics included in the analysis (i.e., the four types of factors discussed above) to predict household gas consumption. The decision tree method is further explained in Section 2.3. This is a promising approach, as a decision tree provides a graphical illustration of which factors are related to household gas use that is relatively simple and easy to understand and interpret. A decision tree method reveals how households with different levels of gas consumption can be characterised, and is useful for exploratory purposes. Notably, a decision tree visualises the relationships between variables in large data sets, and not only identifies the most significant variables related to gas consumption, but also interactions between two or more predictor variables included in the model. This is a major advantage above multiple regression analysis and therefore a decision tree is likely to provide a more comprehensive insight into factors related to household gas use. Moreover, unlike other statistical tools that aim to assess relationships between variables, decision trees remain flexible to handle both continues and categorical variables and variables with some missing variables. Additionally, unlike linear regression and logistic regression models, a decision tree does not require any assumptions of linearity in the data.

2. Method

2.1. Data collection and pre-processing

A questionnaire study was conducted among households having a smart meter and that had been customer of a Dutch energy provider, Qurrent, for at least 6 months, in the Netherlands in 2017. Participants were recruited via an email sent by the energy provider, and were asked to fill out a questionnaire and to share their gas consumption data collected via their smart meter with the research team. One person per household filled out the questionnaire. The study was part of the EU-funded project "Psychological, social and financial barriers to energy efficiency".[1] Data were collected on building characteristics, socio-demographic variables, gas use behaviour and psychological variables. Additionally, household gas use data were collected based on smart meter readings, reflecting bimonthly gas use of the years 2017–2018 (from the first of January 2017 to the first of July 2018). A total of 2318 households completed the questionnaire. However, data on gas use, the dependent variable, were available for 1211 households only. Of those 1211 households, 610 were excluded from data analyses based on the following three criteria:

(1) When households answered "don't know" to the following questions: "how big is your house in term of square meters?", "in which of the following periods was your house originally built?", "which of the following best describe your household type?", "how much was your household's total monthly gross income in the following years?", "what is the usual temperature in your living room during winter at day time and night time in winter?", as these questions tap on important predictors of household gas use. In total 160 cases were excluded for this reason.

(2) When households selected "apartment" as their residence type. This study aimed at focusing on "detached house", "semi-

detached house" and "terraced house" as these type of residences are the most common dwelling types in the Netherlands. Moreover, people living in an apartment completed some different questions than those living in detached house, semi-detached house and terraced house, so including them would have resulted in many missing data on key variables. In total 283 respondents lived in apartments and where therefore excluded.

(3) When data on total gas consumption in the years 2017 and 2018 were missing. In total 167 cases were excluded for this reason.

Hence, in total 601 households were included in the analyses.[2]

2.2. Measures

As explained in the Introduction, variables from four categories of predictor variables were included: building characteristics, socio-demographic variables, gas use behaviour and psychological factors. Tables 1–4 give an overview of all variables included in the analyses.

2.2.1. Building characteristics

Households were asked to indicate the type of dwelling they live in (i. e. detached house, semi-detached house and terraced house), the size of dwelling they live in (in square meter), the number of rooms in the dwelling including habitable spaces such as living room, dining room, bedroom and office, and the year of construction of the dwelling. Table 1 shows the building characteristics that were used and their response percentages.

2.2.2. Socio-demographic variables

Table 2 displays descriptives for the socio-demographic variables. In total 183 females and 418 males participated in the study. Age ranged from 19 to 55 (M = 37.99, SD = 9.37). Respondents indicated the highest educational degree they have completed, their employment status and the total household monthly gross income in year 2016 as well as their household type. Household type included five categories, namely "single person", "single parent with one or more children", "couple, without children", "couple with one or more children" and "other type of household". To facilitate the interpretation of the result, household type was recoded into two dichotomous variables, namely "single" and "couple". Next, for singles, "single without children" (0) and "single parent with one or more children" (1) were distinguished. Similarly, "couple without children" (0) and "couple with one or more children" (1) were distinguished. None of respondents indicated to live in an "other type household".

2.2.3. Gas use behaviour

Respondents were asked to indicate the usual temperature in their

Table 1
Descriptives of building characteristics (% of respondents).

Variable	Response categories (percentage)
House size (m^2)	110 or less (30.7%), 120–150 (35.6%), 160 or more (33.7%)
Number of rooms	1 (0.4%), 2 (2.5%), 3 (12.8%), 4 (36.9%), 5 (26.1%), 6 (21.3%)
Residence type	Detached (23.8%), Semi-detached (23.6%), Terraced (52.6%)
Year built	Before 1940 (26.8%), 1940–1970 (18.3%), 1971–2000 (39%), 2001 or later (15.9%)

---

[2] A power calculation with taking into account 95% confidence level and 5% margin of error (confidence interval) indicates that the sample size of 601 is considered as sufficient to explore how the four types of factors are related to household gas consumption.

Table 2
Descriptives for socio-demographic variables.

Variable	Response categories (percentage)
Respondent age	Continuous: M = 37.99, SD = 9.37, Min = 19, Max = 55
Household type: single	Single with no children (20.6%), Single with children (79.4%),
Household type: couple	Couple, no children (54.4%), Couple with children (45.6%)
Educational degree	Lower secondary school (6.7%), Vocational secondary school (15.9%), Upper secondary school (7.1%), Higher education (42.4%), 5-year university degree (21.4%), Postgraduate qualification (6.5%)
Employment status	Employed, full time (42.7%), Employed, part time (19.2%), Self-employed (10%), Seeking work (2.9%), Student (.3%), House-wife/man (1.8%), Retired (20.6%), Other (2.5%)
Income	Below 1500 (3%), 1500–2500 (10.5%), 2501–3500 (14.3%), 3501–4500 (19.3%), 4501–6000 (20.9%), 6001–9000 (19.3%), 9001–12,000 (6%), Above 12,000 (6.7%)

Table 3
Descriptives for room temperature settings during day and night time, respectively, across resident type (n = 601).

Variable	Day time		Night time	
	M	SD	M	SD
Room temperature	19.42	1.33	16.65	1.72

Table 4
Descriptive for the psychological factors.

Variable	Cronbach's alpha	M (SD)	Min	Max
Biospheric values	.88	4.55 (1.37)	.00	7.00
Egoistic values	.73	5.23 (1.26)	-.75	7.00
Hedonic values	.81	5.17 (1.17)	.25	7.00
Altruistic values	.76	5.15 (1.12)	1.67	7.00
Corporate environmental responsibility	.85	5.53 (1.09)	1.00	7.00
Pro-environmental self-identity	.88	5.48 (1.12)	1.00	7.00
Personal norm	.84	4.56 (1.11)	1.00	7.00
Social norm	.72	5.63 (1.14)	1.00	7.00

living room during day time and night time, respectively, in winter, in degrees Celsius. Table 3 shows descriptives of room temperature settings as an indicator of gas use behaviour. The largest proportion of the sample reported room temperature settings during day time to be 20 °C degree (35.2%), while the room temperature settings during night time was mostly below 16 °C (33.7%).

2.2.4. Psychological factors

Respondents filled in a brief value questionnaire to measure their biospheric, egoistic, hedonic and altruistic values (see Ref. [22]) consisting of 16 items. A brief explanation was given of the relevant values. The scale included four biospheric values (Respecting the earth: harmony with other species; Unity with nature: fitting into nature; Protecting the environment: preserving nature; Preventing pollution: protecting natural resources, four altruistic values (Equality: equal opportunity for all; A world at peace: free of war and conflict; Social justice: correcting injustice, care for the weak; Helpful: working for the welfare of others), five egoistic values (Social power: control over others, dominance; Wealth: material possessions, money; Authority: the right to lead or command; Influential: having an impact on people and

events; Ambitious: hardworking, aspiring), and three hedonic values (Pleasure: joy, gratification of desires; Enjoying life: enjoying food, sex, leisure, etc.; Self-indulgent: doing pleasant things). Respondents were asked to indicate to what extent these values were important to them as a guiding principle of their life, on a 9-point scale (−1 opposed to my values to 0 not important to 7 extremely important). Following Schwartz [20,40], respondents were advised to differentiate as much as possible between the items, and to rate no more than two values as extremely important, to ensure that participants distinguished between the importance of the different values. The items of the biospheric value scale formed a reliable scale[3] α = 0.88 (M = 4.55, SD = 1.37). The same was true for egoistic values (α = 0.73, M = 5.23, SD = 1.26), hedonic values (α = 0.81, M = 5.17, SD = 1.17), and altruistic values (α = 0.76, M = 5.15, SD = 1.12). Thus, all value scales had sufficient internal consistency. The mean scores indicate that people generally rather strongly endorse egoistic, hedonic and altruistic values, while biospheric values were relatively less important to people.

Corporate environmental responsibility was measured with the following three items reflecting the extent to which people think their energy provider has the aim to improve its environmental performance and to reduce its environmental impact (cf [39]): I think that my energy provider has the goal to minimise its impact on the environment; I think that my energy provider has implemented policy and procedures to minimalize its impact on the environment; I think that my energy provider has stated in its mission to implement sustainable (pro-environmental) policy. Respondents indicated to what extent they agree with the items on a 7-point scale ranging from 1 (totally disagree) to 7 (totally agree). Cronbach's alpha for this scale was 0.85 (M = 5.53, SD = 1.09), again indicating that the items formed a reliable scale. The mean score indicates that respondents generally believe that their energy provider has a clear mission to reduce its environmental impact.

A validated scale was used to measure pro-environmental self-identity, comprising three items: Acting pro-environmentally is an important part of who I am; I am the type of person who acts pro-environmentally; I see myself as a pro-environmentally person [27,32]. Respondents rated each item on a 7-point scale, ranging from 1 totally disagree to 7 totally agree. The items of this scale formed a reliable scale as well: α = 0.88 (M = 5.48, SD = 1.12). The mean score indicates that respondents generally see themselves as a person who acts pro-environmentally.

Personal norm to save energy was measured with four items: I feel morally obliged to save energy; It is my moral ideal to save energy; I would act according to my principles if I save energy; I feel personal responsible to try to save energy. Respondents indicated to what extent they agree with the items on a 7-point scale ranging from 1 (totally disagree) to 7 (totally agree). Cronbach's alpha was again good: 0.84 (M = 4.56, SD = 1.11). The mean score indicates that respondents generally do not experience a very strong personal norm to save energy.

Social norm was measured with three items which reflect injunctive norms and descriptive norms: Most of the people who are important to me think I should try to use less energy; Most of the people who are important to me will approve that I try to use less energy; Most people who are important to me try to use less energy. Respondents rated each item on a 7-point scale, ranging from 1 'totally disagree' to 7 'totally agree'. The items of this scale formed a reliable scale as well: α = 0.72 (M = 5.63, SD = 1.14). The mean score indicates that respondents generally experience relatively strong social norm to save energy. Table 4 reports descriptives for the psychological variables.

---

[3] Cronbach's Alpha (α), a measure that reflects to what extent the items in the scale are likely to reflect the same underlying construct, was used to determine whether the items formed reliable (or: internally consistent) value scales. Higher scores of α indicate that the items included in the scale are likely to measure the same underlying construct [41]. The general rule of thumb is that a Cronbach's alpha of .70 are higher is good.

2.2.5. Total gas consumption

Gas consumption data were derived from smart meters. Fig. 1 visualises the distribution of the bimonthly household gas consumption and its probability density between January 2017 and July 2018. It displays the distribution shape of household gas consumption for each month. Wider sections of the violin plot represent a higher probability that households in the sample will use the given amount of gas, while the skinnier sections represent a lower probability that households in the sample have consumed the given amount of gas.

The total household gas consumption from January 1, 2017 to July 1, 2018 in cubic meters (m^3) was included as dependent variable in the analyses. The mean gas consumption in these 1.5 years was 2047.25 (SD = 1465.40). Gas consumption of the participating households in 2017 (M = 1015.68) was very similar to the average gas consumption of Dutch households in 2017 (M = 998.46).[4] Total gas consumption was transformed into its cube root using Box-Cox transformation[5] to achieve normality of the gas consumption distribution. The mean Box-Cox transformed gas consumption was M = 12.01 with a standard deviation of SD = 2.94.

2.3. Statistical analysis: decision tree

To identify the most important factors related to household gas consumption, a decision tree model [42] was performed for the bimonthly household Box-Cox transformed gas consumption (in m^3) between January 2017 and July 2018. Building characteristics, socio-demographic variables, household behaviour and psychological factors were included as predictors of total gas consumption.

Decision tree learning methods involving continuous variables, like in the current study, are called regression trees and date back to Ref. [43]. Although this machine learning approach is widely used in other fields, they are relatively unknown in psychological research and energy research despite their clear conceptual advantages over regression models.

Decision trees are tree shaped diagrams, where through a series of dichotomous classifications the data set is split into a number of subsets. When such classifications are made for continuous predictors, an algorithm finds some optimal threshold value for classifying households based on total gas use: values of the continuous variable below the threshold, and those above the threshold, are classified into two distinct branches. With each sub split, the proportion of unexplained variance is reduced, at the cost of a more complicated model. The final result is a tree with so-called decision nodes and leaf nodes. Each decision node (reflecting a predictor variable) has two branches, defining a binary classification of lower versus higher gas consumption on the basis of the relevant predictor variable. The nodes in the final layer of the tree are denoted leaf nodes. Hence, a decision tree reflects which are the most important predictors of gas consumption, and which values of the predictor variable distinguish households with a lower versus higher gas consumption.

The decision tree reveals the average gas used by households having the specific characteristics (first number below the nodes in Fig. 2), as well as the percentage of households falling into each category (second number below the nodes in Fig. 2). Notably, by following a path from the root note (the top) to a leaf node (the bottom), one obtains a set of decision rules that result in the class of households with the given average gas consumption. Therefore, to understand which factors can explain gas consumption in the decision tree, one needs to go from the top to decision tree via each branch to the bottom, which reflects how households included in each category can be characterised.

In order to check whether the results were robust, the dataset was randomly split into a training set and testing set. Typically, the training set comprises a larger proportion of the sample, while a smaller portion of the data is used for testing. The training set is used to construct the model and the test set is used to validate the model derived on the basis of the training set. The dataset was divided into a training set and a test set, respectively, with the ratio 80:20. After a model has been processed by using the training set, the prediction accuracy of the decision tree model can be evaluated. As the target variable, total gas consumption, is continuous, Mean Absolute Error (MAE[6]) and Root Mean Squared Error (RMSE[7]), which provide a reliable indication of the fitness of the model, are both used to test the accuracy of the model. Specifically, MAE and RMSE are computed for both training and testing set. If the value of MAE and RMSE are very similar for the training set and the test set, then it can be concluded that a good model was built [44,45] In this study, the model was fit and visualised using the rpart and rpart.plot packages [46, 47] in R [48].

3. Results

The decision tree was built on the basis of the training data set which included 487 respondents. Fig. 2 displays the decision tree model to explain gas consumption. It shows the mean gas consumption of households in each branch (first number in each node), and the percentage of households in the sample that end up in each branch (second number in each node). The decision tree has a total of 33 nodes among which 17 are leaf nodes. The results indicate that building characteristics, socio-demographic variables and psychological factors are all important predictors in explaining household gas consumption. Specifically, house size, residence type, and building age were the main building characteristics that are related to household gas use. Income and employment status were the main socio-demographic variables explaining gas use. Egoistic values, hedonic values, environmental self-identity, corporate environmental responsibility and social norm were the main psychological factors that are related to gas consumption. Interestingly, room temperature setting as an indicator of gas use household behaviour was not significantly related to household gas consumption.

House size appears to be the best predictor of gas consumption, as it is the starting point of the decision tree. Specifically, the decision tree splits households into two branches: those living in a house with 160 square meter or larger use more gas than those living in a house smaller than 150 square meter.

For houses larger than 160 square meter, income was the next best predictor of gas consumption: households with an income higher than 12,000 euro use more gas than households with a lower income. In the next step of this branch, for households with a lower income, employment status is the most important predictor: respondents who are either full-time employed, self-employed, seeking work, student, house-wife/man or retired use more gas than respondents who are part-time employed or those have other types of employment. For respondents who are part-time employed or have other types of employment, hedonic values are the next best predictor of gas use: stronger hedonic values (i.e., score higher or equal to 3.3) imply that a household uses less gas than when hedonic values are weak.

For respondents with an employment status of full time employed, self-employed, seeking work, student, house-wife/man or retired, the building year of the house is the next best predictor of gas use: houses built before 1940 are associated with a lower gas consumption than houses built later. For households living in a house built before 1940, corporate environmental responsibility is the next best predictor of gas use: a weaker perceived corporate environmental responsibility (i.e.,

---

[4] Gas use data in 2018 could not be compared with the gas use of Dutch households in 2018 as the latter data were not available yet.

[5] The Box Cox transformation implies that the non-normal total gas consumption distribution was transformed into a normal distribution.

[6] $MAE = \frac{1}{n} \times \sum |Y_{obs} - Y_{pred}|$

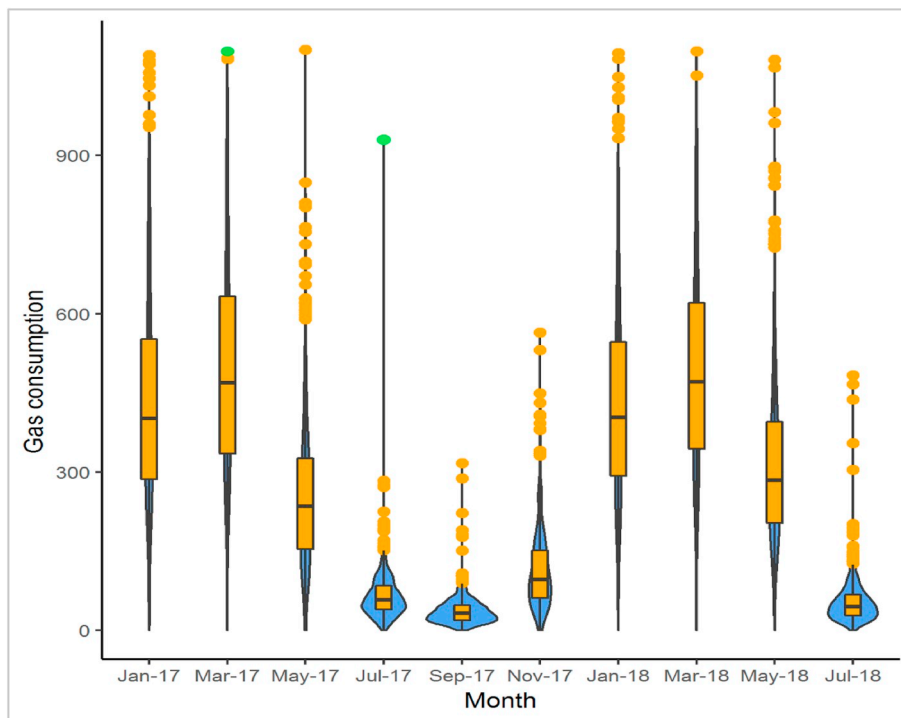[7] $RMSE = \sqrt{\frac{1}{n} \times \sum (Y_{obs} - Y_{pred})^2}$

Fig. 1. Violin plot with boxplot of bimonthly household gas consumption (m^3) ($n = 601$) for the period January 2017- July 2018. The figure is a combination of a boxplot and a density plot to show the distribution shape of the data. The black line in the middle of boxplot is the median value of household gas consumption, and the thick yellow bar in the centre represents the interquartile range. The thin black line extended from it represents the upper (max) and lower (min) adjacent values in the data. The blue part visualises the distribution shape of household gas consumption over the month and its width shows how many households used the given amount of gas. The yellow points represent the outlier of the boxplot for household gas consumption over the month. Two households with a very high gas consumption (one of 1098 m^3 in March 2017, and one of 930 m^3 in July 2017) are printed as green dots the figure. The Figure shows that distributions of household gas consumption are most uniform in July 2017, September 2017 and July 2018. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

score lower than 4.8) is associated with a lower gas consumption, while a stronger perceived corporate environmental responsibility is related to a higher gas consumption. For households living in a house built between 1940 and 1970 or later, employment status is the next most important predictor of gas use: respondents who are either full time employed, house-wife/man or retired use less gas than respondents who are part-time employed, self-employed, seeking work, student, or those having other types of employment. For high income households, environmental self-identity is the next best predictor of gas use: a stronger environmental self-identity (i.e., score higher or equal to 5.5) is associated with a lower gas consumption than a weaker environmental self-identity.

For houses smaller than 150 square meter, residence type was the next best predictor of gas consumption: households living in a terraced houses use less gas, compared to households living in a detached houses or semi-detached houses. In the next step, for households living in a terraced house, employment status is the most important next predictor of gas use: those who are full-time employed, part-time employed, retired, house-wife/man or seeking work use less gas than respondents who are self-employed, student or have other types of employment. For respondents who are full-time employed, part-time employed, retired, house-wife/man or seeking work, egoistic values are the next best predictor of gas use: stronger egoistic values (i.e., score higher or equal to 2.5) are associated with a lower gas consumption than weaker egoistic values. For respondents with relatively strong egoistic value (score 2.5 or higher), social norm is the next best predictor of gas use: respondents who report a weaker social norm to save energy (i.e., score lower than 5.8) use less gas than those who report a stronger social norm. For respondents with weaker egoistic values, hedonic values are the next predictor: stronger hedonic values (i.e., score higher or equal to 4.8) are associated with a lower gas consumption than weaker hedonic values. For households living in a detached houses or semi-detached houses, income is the next best predictor of gas use: households with an income lower than 4500 euro use less gas than households with an income of 4500 euro or higher. For households with a lower income, egoistic values are the next best predictor of gas use: stronger egoistic values (i.e., score higher or equal to 2.1) are associated with a higher gas

consumption than weaker egoistic values. For households with a higher income, the building year of the house is the next best predictor: households living in a house built between 1971 and 2000 have a higher gas consumption than households living in a house built in 2001 or later.

Notably, the households with the highest gas consumption (mean transformed gas use of 15.62 m^3) live in a house of 160 square meter or more, their monthly gross income is less than 12,000 euro, they are full-time employed, self-employed, seeking work, student, house-wife/man or retired, their house is built between 1940 and 1970 or later and they are part-time employed, self-employed, seeking work, student or have other types of employment; they represent 2% of the households in the sample. Households with the second highest gas consumption (mean transformed gas use of 15.48 m^3) live in a house of 160 square meter or more, their monthly gross income is more than 12,000 euro, and they have a weaker environmental self-identity (i.e., score lower than 5.5); they represent 4% of the households in the sample. The group of households with the lowest gas consumption (mean transformed gas use of 8.9 m^3) live in a house that is 160 square meter or larger, their income is less than 12,000 euro, they are full-time employed, self-employed, seeking work, student, house-wife/man or retired, their house is built before 1940 and they believe their energy provider does not strongly endorse corporate environmental responsibility (i.e., score lower than 4.8); only 1% of the households of the sample ended up in this branch. Appendix A provides a full overview of the decision rules derived from the obtained decision tree, resulting in different classes of gas consumption.

The result further showed that for the training set MAE and RMSE were 1.99 and 2.52, respectively. The MAE for the test set was 2.1, while RMSE was 2.76. As the MAE and RMSE are very similar for the training set and the test set, it can be concluded that the resulting decision tree model is accurate [45].

Finally, a multiple regression analysis was conducted for the same data set and these results were compared with the decision tree model reported above. The regression model explained 20% of the variance in cube root-transformed household gas consumption (adjusted $R^2 = 14.7\%$, $F (31, 455) = 3.707$, $p < .001$); residence type and income were the strongest predictors. Table 5 shows that the regression analysis
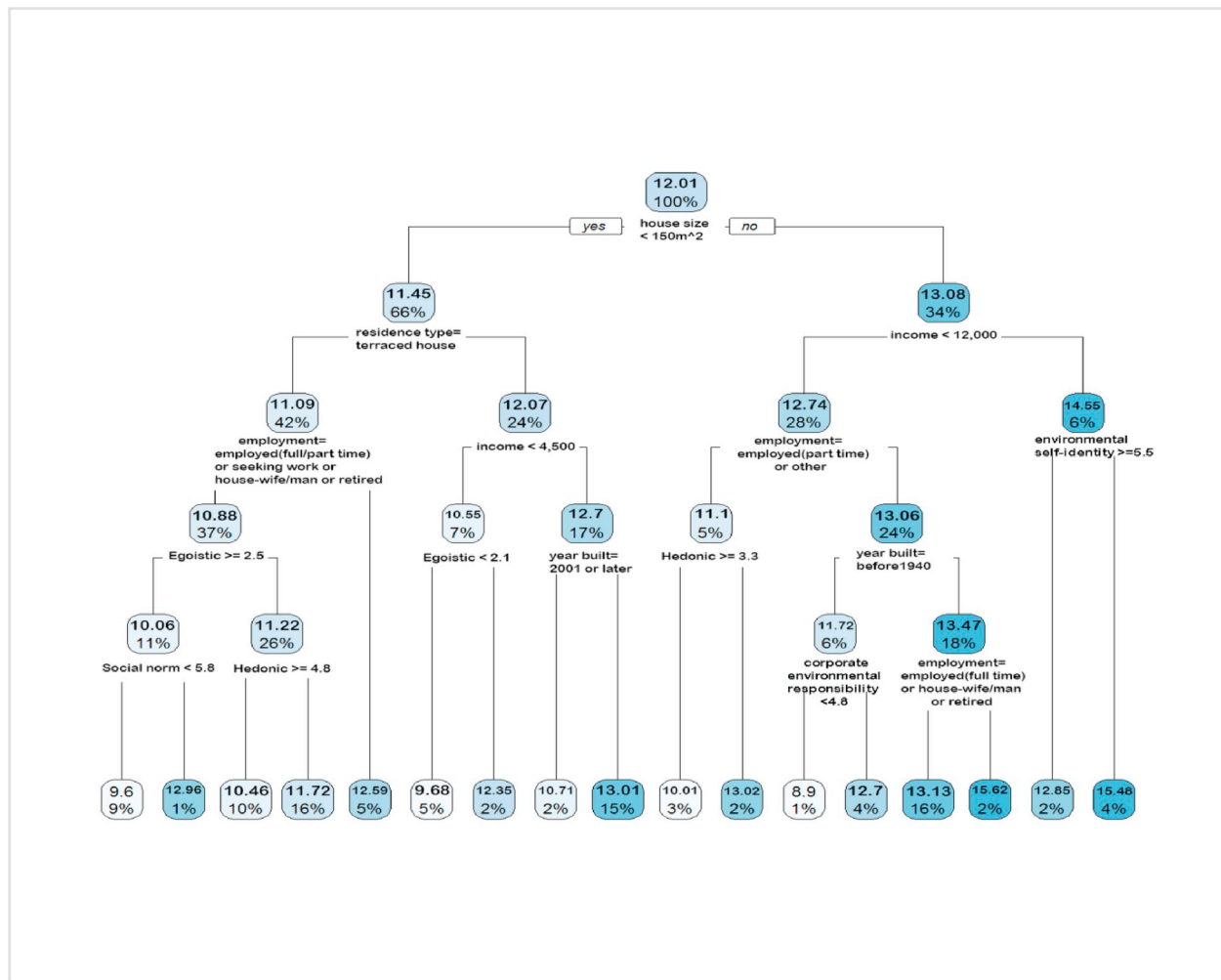
Fig. 2. Decision tree for explaining total gas use in Dutch households. The first number in each node box represents the average gas consumption of households in this branch, while the second number in each node box indicates the percentage of households in the sample that end up in this branch. The colour of the node boxes reflects the fitted value of the average gas consumption of households in each branch: small fitted values are displayed with a lighter hue of blue and large values are shown with darker shade of blue. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

revealed that households living in terraced houses use less gas than households living in a detached house, while higher income groups use more gas. A larger number of rooms in the residence and a weaker environmental self-identity were related to a higher gas consumption too. Self-employed respondents use more gas than those who are employed-full time, while those with a postgraduate qualification use less gas than those with a lower secondary school degree. The results are partly similar to the results of the decision tree, with residence type, income, employment status and environmental self-identity being significant predictors in both models. Number of rooms and educational degree were only significant predictors of gas use in the regression analysis. Yet, the decision tree model yielded a more comprehensive picture of relevant predictors of household gas use, and also identified house size, building age, egoistic values, hedonic values, corporate environmental responsibility and social norm as relevant predictors.

4. Discussion

This paper examined whether gas energy consumption of Dutch households could be explained by building characteristics, socio-demographic variables, household behaviour and psychological variables using a novel method, a decision tree model. Extending previous research that typically studied the predictive power of these types of categories separately [[6–8,10,14,19,49–51]], the results showed that

building characteristics, socio-demographic and psychological variables are all significant and unique predictors of household gas consumption. This is an important and novel finding that clearly signals that an integrated approach is needed to better understand household gas consumption, as taking into account only one type of predictor will provide a limited understanding of household gas consumption.

In terms of building characteristics, particularly house size, and to a lesser degree residence type and building age were important predictors of household gas consumption. These findings are in line with earlier studies that revealed that these building characteristics are related to gas consumption [[5–8,10,12]]. House size was the best predictor of household gas consumption. Not surprisingly, larger dwellings (i.e., 160 square meter or more) are more likely to have a higher gas consumption. Furthermore, households living in terraced houses use less gas, compared to households living in a detached houses or semi-detached houses. These findings are in line with previous studies that have found that those living in detached house and semi-detached house used more gas than those living in other type of dwellings [5,6]. This can probably be explained by the fact that detached houses and semi-detached houses have more external walls and a larger outside wall area, and therefore have more heat losses, resulting in a higher gas consumption. Building age was the least important building characteristic predicting gas use. Specially, households living in larger houses, with a lower income, of which the respondent is either full-time

Table 5

Results for the multiple regression analysis of the transformed total actual gas use (m^3) of Dutch households, including building characteristics, socio-demographic variables, gas use behaviour and psychological factors as predictor variables.

Variable	B	SE	t	p-value	95% CI
Intercept	9.913	1.729	5.732	<.001***	(6.514; 13.311)
House size (m^2)	0.302	.188	1.601	.110	(-.068; .673)
Number of rooms	.275	.135	2.028	.043*	(.008; .542)
Residence type (Ref = detached)					
Residence type = semi-detached	-.745	.385	−1.931	.054	(-1.503; .013)
Residence type = terraced	−1.443	.367	−3.932	<.001***	(-2.164; −.722)
Year built	-.200	.130	−1.541	.124	(-.456; .055)
Respondent age	-.017	.014	−1.192	.234	(-.045; .011)
Household type = single	.449	.289	1.555	.120	(-.118; 1.017)
Household type = couple	-.139	.355	-.392	.695	(-.838; .559)
Educational degree (Ref = lower secondary school)					
Educational degree = vocational secondary school	−1.138	.622	−1.828	.068	(-2.362; .085)
Educational degree = upper secondary school	−1.185	.715	−1.657	.098	(-2.590; .220)
Educational degree = higher education	-.551	.575	-.959	.337	(-1.682; .578)
Educational degree = 5-year university degree	−1.227	.632	−1.940	.052	(-2.470; .015)
Educational degree = postgraduate qualification	−1.511	.766	−1.971	.049*	(-3.018; −.004)
Employment status (Ref = employed, full time)					
Employment status = employed, part time	.028	.360	.078	.938	(-.680; .736)
Employment status = self-employed	.985	.469	2.100	.036*	(.063; 1.908)
Employment status = seeking work	.704	.745	.945	.345	(-.760; 2.169)
Employment status = student	.704	.745	.945	.345	(-.621; 10.613)
Employment status = house-wife/man	.650	1.109	.586	.558	(-1.529; 2.829)
Employment status = retired	.650	1.109	.586	.558	(-.509; 1.099)
Employment status = other	-.141	.918	-.154	.877	(-1.947; 1.664)
Income	.294	.092	3.190	<.001***	(.113; .476)
Room temperature (day time)	.179	.114	1.564	.118	(-.045; .404)
Room temperature (night time)	-.094	.086	−1.090	.276	(.265; .076)
Biospheric values	.081	.163	.502	.616	(-.238; .402)
Egoistic values	.081	.122	.665	.506	(-.158; .321)
Hedonic values	-.155	.110	−1.404	.160	(-.372; .061)
Altruistic values	.141	.141	1.003	.316	(-.135; .418)
	.190	.123	1.536	.125	

Table 5 (*continued*)

Variable	B	SE	t	p-value	95% CI
Corporate environmental responsibility					(-.053; .433)
Pro-environmental self-identity	-.366	.185	−1.981	.048*	(-.730; −.002)
Personal norm	-.017	.188	-.093	.926	(-.388; .353)
Social norm	.042	.140	.303	.762	(-.233; .318)

*p < .05; **p < .01; ***p < .001.

employed, self-employed, seeking work, student, house-wife/man or retired, and live in an older houses use less gas than the same group that live in a newer houses. However, households living in a smaller house, living in a detached house or semi-detached house, with a higher income and living in newer houses use less gas than those living in older houses. Interestingly, number of rooms in the dwelling was not a significant predictor of household gas consumption. Importantly, extending previous research, the results show that these building characteristics uniquely predict household gas consumption, when all other variables are included in the model as well.

Of the socio-demographic variables, income and employment status were the main predictors of household gas consumption. Income was a relatively important predictor of gas consumption for households living in a larger house (i.e. 160 square meter or larger): households living in larger houses with a higher income use more gas, suggesting that people with a higher income seem to particularly use more gas when they also live in larger dwellings. These findings are in line with previous studies [11,13,14], but extend it by showing that income is also related to household gas consumption when the other predictors are included in the model too. Interestingly, respondent age, household type and education level did not significantly contribute to the explanation of household gas consumption when the other variables were controlled for.

The most notable and novel result of this study is that psychological factors play an important role in explaining household gas consumption. Specifically, egoistic values, hedonic values, environmental self-identity, perceived corporate environmental responsibility of the energy provider and social norm to reduce energy use were all uniquely and significantly related to household gas consumption. Interestingly, households living in larger dwellings, with a higher income and with a weaker environmental self-identity are more likely to be in a category with the second highest level of gas consumption. It may be that people living in larger houses with a higher income are less likely to see themselves as the type of persons who engage in pro-environmental actions, which may imply that they are less motivated to act pro-environmentally, which could explain their higher gas use. Moreover, people living in a smaller house, and a detached house or semi-detached house, with a lower income and stronger egoistic values are more likely to use more gas than the same group with weaker egoistic values. This is in line with earlier studies that revealed that egoistic values are often negatively related to pro-environmental behaviours [25]. However, households living in a smaller house, and a terraced house, who are either full-time employed, part-time employed, retired, house-wife/man or seeking work and have stronger egoistic values use less gas than the same group with weaker egoistic values. This is contrary to what would be expected. This may be due to the fact that this finding is specific to a subsample as reflected in this branch of the decision tree. Another explanation could be that respondents in this branch of the decision tree cannot afford to use more gas, or are forced by the circumstances to use less gas. Future research is needed to explore the negative relation of egoistic values with household gas consumption and why some groups with stronger egoistic values use less gas.

Interestingly, this study showed that perceived corporate environmental responsibility of the utility company can explain household gas

consumption. This is a novel and interesting finding indicating that households living in larger houses, with a lower income, who are either full-time employed, self-employed, seeking work, student, house-wife/man or retired, living in a house that is built before 1940 and think their energy provider does not strongly endorse corporate environmental responsibility have the lowest gas consumption. Yet, only a very small group of households (1%) of the sample falls into this branch. This finding is contrary to the expectations, as a negative relation was expected between corporate environmental responsibility and gas consumption, with a stronger corporate environmental responsibility being associated with using less gas. To explore this further, the bivariate correlation was inspected between corporate environmental responsibility and gas use for the whole sample of the training set that revealed a similar positive relation between corporate environmental responsibility and gas consumption. Future research is needed to examine why a stronger perception of corporate environmental responsibility of the energy provider is related to using more gas.

In contrast to the expectations, it was found that people living in a smaller house, and a terraced house, who are either full-time employed, part-time employed, retired, house-wife/man or seeking work, with stronger egoistic values and weaker social norm to save energy use less gas than the same group with a stronger social norm to save energy. Besides, people living in a larger house, and with a lower income, who are part-time employed or have other types of employment and with stronger hedonic values use less gas than the same group with weaker hedonic values. Moreover, households living in a smaller house, and a terraced house, who are either full-time employed, part-time employed, retired, house-wife/man or seeking work, with weaker egoistic values and with stronger hedonic values use less gas than the same group with weaker hedonic values. Future research could explore why these phenomena occur.

In sum, extending previous research, these results indicate that psychological variables uniquely explain household gas consumption, next to building characteristics and socio-demographics. These findings are interesting and in line with earlier studies that reveal that psychological factors play an important role in explaining household energy use [52], and extend this research by showing that psychological factors are also important to understand household gas use.

Interestingly, indoor temperature settings during day time and night time did not appear in the decision tree, and is thus not uniquely associated to household gas consumption. This may be explained by findings from earlier results that show that room temperature setting can be predicted by socio-demographic variables, environmental values and building characteristics [53]. These factors were also identified as predictors of household gas consumption in the study. Thus, perhaps for this reason, indoor temperature setting room during day time and night time does not explain unique variance in household gas consumption. Table 6 provides a summary overview of the factors that appeared to be significantly related to household gas consumption.

This study aimed at exploring which factors are associated with actual household gas consumption using a decision tree model for selecting key determinants. The results show that the decision tree is an appropriate method for exploratory analysis by detecting important variables related to household gas consumption, and visualising the relationships between different predictor variables. One of the assets of decision tree method is identifying possible interaction effect between the predictors. This study revealed several novel and interesting interactions between variables predicting gas use, such as households living in (semi-) detached houses having a higher gas consumption when they live in larger dwelling size. Likewise, people with a higher income seem to particularly use more when they also live in larger dwellings.

Furthermore, the decision tree classified households on the basis of their total actual gas consumption, and revealed that different numbers of predictors were needed for identifying different classes of households differing in gas consumption. Indeed, fewer predictors were needed for explaining household gas consumption in some branches, whereas in

Table 6
Summary of factors that are significantly related to household gas consumption.

Categories	Specific factors	Significant predictors
Building characteristics	House size	✓
	Number of rooms	
	Residence type	✓
	Year built	✓
Socio-demographic variables	Respondent age	
	Household type	
	Educational degree	
	Employment status	✓
	Gross household income	✓
Gas use behaviour	Room temperature setting during daytime	
	Room temperature setting during night time	
Psychological factors	Biospheric values	
	Egoistic values	✓
	Hedonic values	✓
	Altruistic values	
	Corporate environmental responsibility	✓
	Pro-environmental self-identity	✓
	Personal norm	
	Social norm	✓

other branches more predictors were needed to explain gas use. Therefore, with specific combination of these predictors, an accurate insight is gained into actual household gas consumption, which would not be attained through standard regression model.

The decision tree model provided novel insights about interactions between predictors, and identified different predictors for different classes of gas consumption. Notably, the decision tree model identified more relevant predictors of household gas use than the multiple regression analysis did. Specifically, the multiple regression analysis including the same predictor and dependent variable revealed that the most important predictors of household gas consumption were residence type and income, while number of rooms in the residence, educational degree, employment status and environmental self-identity were also significantly but weakly related to household gas consumption. Hence, the decision tree model yielded more comprehensive and sophisticated insights into relevant predictors of household gas use than the multiple regression analysis, by also identifying house size, building age, egoistic values, hedonic values, corporate environmental responsibility and social norm as relevant predictors (while number of rooms and educational degree were significant predictors in the multiple regression analysis, but not in the decision tree model). Interestingly, the decision tree model particularly identified more psychological variables as relevant predictors of gas use than the regression model did. The decision tree model thus provided more nuanced and richer insights into which variables could best be targeted to encourage households to reduce their gas use and therefore become more sustainable.

Although this study included a wider range of factors that may explain household gas use than most earlier studies on household gas use, still other factors may be relevant to understand household gas use. Future studies could study to what extent other building characteristics (e.g. level of insulation), household characteristics (e.g. amount of time household members are present in the home), psychological variables (e.g. concern about climate change), and occupant behaviour (e.g. number of heated rooms) would be uniquely related to household gas consumption. Furthermore, the sample of this study only included Dutch households who are a client of a specific energy company (i.e., Qurrent). Future research could examine whether the results of this study would be replicated among other samples, including general population samples.

In this study, household gas use data were collected via smart meter readings which is a far more accurate assessment of gas consumption than self-reports that are oftentimes used in research. As such, the findings reported in this paper provide relevant practitioners, such as utility companies and consumers, with better insights into predictors of actual household gas consumption.

The results of this study have important practical implications, and can support relevant stakeholders, including governments, energy supplier and companies in developing interventions that are aimed at reducing household gas use. The result suggest that policy aimed to reduce household gas consumption can best target building characteristics, socio-demographic variables and psychological factors. Specifically, interventions could particularly consider house size, residence type, building age, income, employment status, egoistic values, hedonic values, environmental self-identity, corporate environmental responsibility and social norm as these appeared to be the main factors that are related to household gas consumption. Particularly, interventions could try to change these predictors, or target groups having the relevant characteristics. For example, information can be provided about the extent to which others find reductions in gas consumption important or that many others try to reduce their gas use, or environmental self-identity can be strengthened, for example by making people aware of their previous sustainable actions [54]. Alternatively, interventions could best target high income groups or people living in large or (semi-) detached houses, as they have a relatively higher gas consumption. As such, the knowledge gained from the analyses reported in this paper support energy policy making by relevant agents aimed at reducing gas consumption in important ways, resulting in lower greenhouse gas emissions. Hence, the information and implications derived from this study provide essential insights into the design of energy policy that can promote reductions in gas consumption as policy strategies are more effective when they target key antecedents of gas use.

5. Conclusions

In this paper, a novel approach, a decision tree method was used to explain household gas consumption. The results show that household gas consumption was uniquely related to buildings characteristics, socio-demographics and psychological factors. Specifically, house size, residence type, and building age were the main building characteristics that are related to household gas use. Income and employment status were the main socio-demographic variables related to gas use. Notably, egoistic values, hedonic values, environmental self-identity, corporate environmental responsibility and social norm were the main psychological variables that are uniquely related to gas consumption. Hence, in order to get a comprehensive understanding of gas use, it is important to consider building characteristics, socio-demographic variables and psychological factors, as they all predict unique variance in household gas use.

Appendix A. Decision rules derived from the obtained decision tree

Leaf nodes	Decision rules associated with different levels of gas use of households
1	If house size $\geq 150 m^2$ and income ≥ 12.000 and environmental self-identity <5.5 then gas consumption is 15.48
2	If house size $\geq 150 m^2$ and income ≥ 12.000 and environmental self-identity ≥ 5.5 then gas consumption is 12.85
3	If house size $\geq 150 m^2$ and income <12.000 and employment status is either full-time employed, self-employed, seeking work, student, house-wife/man or retired and year built and employment status is either part-time employed, self-employed, seeking work, student, or other types of employment then gas consumption is 15.62
4	If house size $\geq 150 m^2$ and income <12.000 and employment status is either full-time employed, self-employed, seeking work, student, house-wife/man or retired and year built ≥ 1940 and employment status is either full time employed, house-wife/man or retired then gas consumption is 13.13
5	If house size $\geq 150 m^2$ and income <12.000 and employment status is either full-time employed, self-employed, seeking work, student, house-wife/man or retired and year built <1940 and corporate environmental responsibility ≥ 4.8 then gas consumption is 12.7
6	If house size $\geq 150 m^2$ and income <12.000 and employment status is either full-time employed, self-employed, seeking work, student, house-wife/man or retired and year built <1940 and corporate environmental responsibility <4.8 then gas consumption is 8.9
7	If house size $\geq 150 m^2$ and income <12.000 and employment status is part-time employed or other types and hedonic <3.3 then gas consumption is 13.02
8	If house size $\geq 150 m^2$ and income <12.000 and employment status is part-time employed or other types and hedonic ≥ 3.3 then gas consumption is 10.01
9	If house size $<150 m^2$ and residence type is detached or semi-detached house and income ≥ 4.500 and year built <2001 then gas consumption is 13.01
10	If house size $<150 m^2$ and residence type is detached or semi-detached house and income ≥ 4.500 and year built ≥ 2001 then gas consumption is 10.71
11	If house size $<150 m^2$ and residence type is detached or semi-detached house and income <4.500 and egoistic value ≥ 2.1 then gas consumption is 12.35
12	If house size $<150 m^2$ and residence type is detached or semi-detached house and income <4.500 and egoistic value < 2.1 then gas consumption is 9.68
13	If house size $<150 m^2$ and residence type is terraced house and employment status is self-employed, student or other types then gas consumption is 12.59
14	If house size $<150 m^2$ and residence type is terraced house and employment status is either full-time employed, part-time employed, retired, house-wife/man or seeking work then egoistic value < 2.5 and hedonic value < 4.8 then gas consumption is 11.72
15	If house size $<150 m^2$ and residence type is terraced house and employment status is either full-time employed, part-time employed, retired, house-wife/man or seeking work then egoistic value < 2.5 and hedonic value ≥ 4.8 then gas consumption is 10.46
16	If house size $<150 m^2$ and residence type is terraced house and employment status is either full-time employed, part-time employed, retired, house-wife/man or seeking work then egoistic value ≥ 2.5 and social norm ≥ 5.8 then gas consumption is 12.96
17	If house size $<150 m^2$ and residence type is terraced house and employment status is either full-time employed, part-time employed, retired, house-wife/man or seeking work then egoistic value ≥ 2.5 and social norm <5.8 then gas consumption is 9.6

References

[1] CBS. Energy - figures - economy | Trends in The Netherlands. https://longreads.cbs.nl/trends18-eng/economy/figures/energy. [Accessed 23 March 2018].

[2] Netherlands S. CO2 emissions virtually unchanged in Q3 2018. Statistics Netherlands. https://www.cbs.nl/en-gb/nieuws/2018/46/co2-emissions-virtually-unchanged-in-q3-2018. [Accessed 12 February 2019].

[3] Energieverbruik door huishoudens. 1990-2018 | Compendium voor de Leefomgeving. https://www.clo.nl/indicatoren/nl0035-energieverbruik-door-de-huishoudens. [Accessed 8 August 2019].

[4] Seyedzadeh S, Rahimian FP, Glesk I, Roper M. Machine learning for estimation of building energy consumption and performance: a review. Visualization in Engineering 2018 Dec 1;6(1):5. https://doi.org/10.1186/s40327-018-0064-7.

[5] Santin OG, Itard L, Visscher H. The effect of occupancy and building characteristics on energy use for space and water heating in Dutch residential stock. Energy Build 2009 Nov 1;41(11):1223–32. https://doi.org/10.1016/j.enbuild.2009.07.002.

[6] Mastrucci A, Baume O, Stazi F, Leopold U. Estimating energy savings for the residential building stock of an entire city: a GIS-based statistical downscaling approach applied to Rotterdam. Energy Build 2014 Jun 1;75:358–67. https://doi.org/10.1016/j.enbuild.2014.02.032.

[7] Harold J, Lyons S, Cullinan J. The determinants of residential gas demand in Ireland. Energy Econ 2015 Sep 1;51:475–83.

[8] Majcen D, Itard L, Visscher H. Statistical model of the heating prediction gap in Dutch dwellings: relative importance of building, household and behavioural characteristics. Energy Build 2015 Oct 15;105:43–59. https://doi.org/10.1016/j.enbuild.2015.07.009.

[9] Brounen D, Kok N, Quigley JM. Residential energy use and conservation: economics and demographics. Eur Econ Rev 2012 Jul 1;56(5):931–45. https://doi.org/10.1016/j.euroecorev.2012.02.007.

[10] Steemers K, Yun GY. Household energy consumption: a study of the role of occupants. Build Res Inf 2009 Nov 1;37(5–6):625–37. https://doi.org/10.1080/09613210903186661.

[11] Mashhoodi B, van Timmeren A. Local determinants of household gas and electricity consumption in Randstad region, Netherlands: application of geographically weighted regression. Spatial Information Research 2018 Dec 1;26 (6):607–18.

[12] Baker KJ, Rylatt RM. Improving the prediction of UK domestic energy-demand using annual consumption-data. Appl Energy 2008 Jun 1;85(6):475–82. https://doi.org/10.1016/j.apenergy.2007.09.004.

[13] Desalu OO, Ojo OO, Ariyibi EK, Kolawole TF, Ogunleye AI. A community survey of the pattern and determinants of household sources of energy for cooking in rural and urban south western, Nigeria. Pan African Medical Journal 2012;12(1).

[14] Hara K, Uwasu M, Kishita Y, Takeda H. Determinant factors of residential consumption and perception of energy conservation: time-series analysis by large-scale questionnaire in Suita, Japan. Energy Policy 2015 Dec 1;87:240–9. https://doi.org/10.1016/j.enpol.2015.09.016.

[15] Van den Brom P, Meijer A, Visscher H. Performance gaps in energy consumption: household groups and building characteristics. Build Res Inf 2018 Jan 2;46(1):54–70. https://doi.org/10.1080/09613218.2017.1312897.

[16] Moll HC, Noorman KJ, Kok R, Engström R, Throne-Holst H, Clark C. Pursuing more sustainable consumption by analyzing household metabolism in European countries and cities. J Ind Ecol 2005 Jan;9(1-2):259–75. https://doi.org/10.1162/1088198054084662.

[17] Chen J, Wang X, Steemers K. A statistical analysis of a residential energy consumption survey study in Hangzhou, China. Energy Build 2013 Nov 1;66:193–202. https://doi.org/10.1016/j.enbuild.2013.07.045.

[18] Boukarta S, Berezowska-Azzag E. Assessing households' gas and electricity consumption: a case study of djelfa, Algeria. Quaest Geogr 2018 Dec 1;37(4):111–29. https://doi.org/10.2478/quageo-2018-0034.

[19] Saeed KH, Sharawi HA. Factors affecting fuel consumption in household cooking in el-salam locality, south kordofan state, Sudan. In: The 5th annual conference – agricultural and veterinary research; 2015.

[20] Schwartz SH. Universals in the content and structure of values: theoretical advances and empirical tests in 20 countries. InAdvances in experimental social psychology 1992 Jan 1;25:1–65. https://doi.org/10.1016/S0065-2601(08)60281-6. Academic Press.

[21] De Groot JI, Steg L. Value orientations to explain beliefs related to environmental significant behavior: how to measure egoistic, altruistic, and biospheric value orientations. Environ Behav 2008 May;40(3):330–54. https://doi.org/10.1177/0013916506297831.

[22] Steg L, Perlaviciute G, Van der Werff E, Lurvink J. The significance of hedonic values for environmentally relevant attitudes, preferences, and actions. Environ Behav 2014 Feb;46(2):163–92. https://doi.org/10.1177/0013916512454730.

[23] Stern PC, Dietz T, Guagnano GA. A brief inventory of values. Educ Psychol Meas 1998 Dec;58(6):984–1001. https://doi.org/10.1177/0013164498058006008.

[24] Boer D, Fischer R. How and when do personal values guide our attitudes and sociality? Explaining cross-cultural variability in attitude–value linkages. Psychol Bull 2013 Sep;139(5):1113. https://doi.org/10.1037/a0031347.

[25] Steg L, de Groot JI. Environmental values. In: The Oxford handbook of environmental and conservation psychology; 2012.

[26] Stern PC, Dietz T. The value basis of environmental concern. J Soc Issues 1994 Oct 1;50(3):65–84. https://doi.org/10.1111/j.1540-4560.1994.tb02420.

[27] Van der Werff E, Steg L, Keizer K. The value of environmental self-identity: the relationship between biospheric values, environmental self-identity and

[28] Lacasse K. Don't be satisfied, identify! Strengthening positive spillover by connecting pro-environmental behaviors to an "environmentalist" label. J Environ Psychol 2016 Dec 1;48:149–58. https://doi.org/10.1016/j.jenvp.2016.09.006.

[29] Uhl I, Jonas E, Klackl J. When climate change information causes undesirable side effects: the influence of environmental self-identity and biospheric values on threat responses/Cuando la información sobre el cambio climático tiene efectos indeseados: la influencia de la identidad ambiental y de los valores biosféricos en la respuesta ante una amenaza. Psyecology 2016 Sep 1;7(3):307–34. https://doi.org/10.1080/21711976.2016.1242228.

[30] Gatersleben B, White E, Abrahamse W, Jackson T, Uzzell D. Values and sustainable lifestyles. InTransforming Markets in the Built Environment 2012 Aug 6:37–50. Routledge.

[31] Whitmarsh L, O'Neill S. Green identity, green living? The role of pro-environmental self-identity in determining consistency across diverse pro-environmental behaviours. J Environ Psychol 2010 Sep 1;30(3):305–14. https://doi.org/10.1016/j.jenvp.2010.01.003.

[32] Van der Werff E, Steg L, Keizer K. It is a moral issue: the relationship between environmental self-identity, obligation-based intrinsic motivation and pro-environmental behaviour. Glob Environ Chang 2013 Oct 1;23(5):1258–65. https://doi.org/10.1016/j.gloenvcha.2013.07.018.

[33] Schwartz SH. Normative influences on altruism. InAdvances in experimental social psychology 1977 Jan 1;10:221–79. https://doi.org/10.1016/S0065-2601(08)60358-5. Academic Press.

[34] Nordlund AM, Garvill J. Effects of values, problem awareness, and personal norm on willingness to reduce personal car use. J Environ Psychol 2003 Dec 1;23(4):339–47. https://doi.org/10.1016/S0272-4944(03)00037-9.

[35] Thøgersen J, Grønhøj A. Electricity saving in households—a social cognitive approach. Energy Policy 2010 Dec 1;38(12):7732–43. https://doi.org/10.1016/j.enpol.2010.08.025.

[36] Schultz PW, Nolan JM, Cialdini RB, Goldstein NJ, Griskevicius V. The constructive, destructive, and reconstructive power of social norms. Psychol Sci 2007 May;18(5):429–34. https://doi.org/10.1111/j.1467-9280.2007.01917.x.

[37] Nolan JM, Schultz PW, Cialdini RB, Goldstein NJ, Griskevicius V. Normative social influence is underdetected. Personal Soc Psychol Bull 2008 Jul;34(7):913–23. https://doi.org/10.1177/0146167208316691.

[38] Cialdini RB, Kallgren CA, Reno RR. A focus theory of normative conduct: a theoretical refinement and reevaluation of the role of norms in human behavior. InAdvances in experimental social psychology 1991 Jan 1;24:201–34. https://doi.org/10.1016/S0065-2601(08)60330-5. Academic Press.

[39] Ruepert AM, Keizer K, Steg L. The relationship between Corporate Environmental Responsibility, employees' biospheric values and pro-environmental behaviour at work. J Environ Psychol 2017 Dec 1;54:65–78. https://doi.org/10.1016/j.jenvp.2017.10.006.

[40] Schwartz S. Value priorities and behavior: Applying. In: The psychology of values: the Ontario symposium, vol. 8; 2013 May 13.

[41] Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychometrika 1951 Sep 1;16(3):297–334.

[42] Quinlan JR. C4. 5: programs for machine learning. Morgan kaufmann, san francisco. C4. 5. San Francisco: Programs for machine learning. Morgan Kaufmann; 1993.

[43] Breiman L. Classification and regression trees. Routledge; 2017 Oct 19.

[44] Gramatica P. Principles of QSAR models validation: internal and external. QSAR Comb Sci 2007 May;26(5):694–701. https://doi.org/10.1002/qsar.200610151.

[45] Guha R, Serra JR, Jurs PC. Generation of QSAR sets with a self-organizing map. J Mol Graph Model 2004 Sep 1;23(1):1–14. https://doi.org/10.1016/j.jmgm.2004.03.003.

[46] Therneau T, Atkinson B, Ripley B, Ripley MB. Package 'rpart'. 2015. cran. ma. ic. uk/web/packages/rpart/rpart.pdf, Jun 29, . [Accessed 20 April 2016].

[47] Milborrow S. Plot 'rpart'models: an enhanced version of 'plot. rpart'. R package version. 2016;2(1).

[48] Team RC. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2014. 2018:2013, https://www.R-project.org/.

[49] Azam M, Khan AQ, Zaman K, Ahmad M. Factors determining energy consumption: evidence from Indonesia, Malaysia and Thailand. Renew Sustain Energy Rev 2015 Feb 1;42:1123–31. https://doi.org/10.1016/j.rser.2014.10.061.

[50] Hsu D. Identifying key variables and interactions in statistical models of building energy consumption using regularization. Energy 2015 Apr 1;83:144–55. https://doi.org/10.1016/j.energy.2015.02.008.

[51] Jeong J, Kim CS, Lee J. Household electricity and gas consumption for heating homes. Energy Policy 2011 May 1;39(5):2679–87. https://doi.org/10.1016/j.enpol.2011.02.037.

[52] Abrahamse W, Steg L. Factors related to household energy use and intention to reduce it: the role of psychological and socio-demographic variables. Hum Ecol Rev 2011 Jun 1;18(1):30–40.

[53] Namazkhan M, Albers C, Steg L. The role of environmental values, socio-demographics and building characteristics in setting room temperatures in winter. Energy 2019 Mar 15;171:1183–92. https://doi.org/10.1016/j.energy.2019.01.113.

[54] Van der Werff E, Steg L, Keizer K. I am what I am, by looking past the present: the influence of biospheric values and past behavior on environmental self-identity. Environ Behav 2014 Jun;46(5):626–57. https://doi.org/10.1177/0013916512475209.